SAS Institute, Ltd. United Kingdom 30 October 2000

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS® System



Andrew H. Karp
Sierra Information Services, Inc.
19229 Sonoma Highway PMB 264
Sonoma, California 95476 USA
707 996 7380
www.SierraInformation.com
SierraInfo@AOL.com

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. ® indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- This presentation will discuss many of the important changes to PROCs MEANS and SUMMARY in Version 8 of the SAS System.
- · You will also see:
 - Some new features of PROC PRINT
 - Use of the Output Delivery System (ODS)

- Both procedures are key elements of BASE SAS Software
 - Very powerful: rapidly summarize and analyze numeric variables in SAS data sets
 - Became identical procedures in Version 6
 - Create:
 - Output SAS data sets (default in PROC SUMMARY)
 - Output in the SAS Output Window (default in PROC MEANS)

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- A basic example of using PROC MEANS to create a SAS Data Set, then we will look at new features in V8
- Task: from a customer order file data set, obtain the total (sum) dollars of sales by mail code and shipment status

- Example Data Set:
 - Customer order file
 - 335,992 observations
 - Potential Classification Variables:
 - Mailcode
 - Mailing Code (Nine values)
 - Dept_Nbr
 - Department number for item (about 40 unique values)
 - Segment
 - Customer demographic segment (six values)
 - Status
 - 1= Shipped Immediately 2 = Deferred Shipment

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Example Data Set (continued):
 - Customer order file
 - 335,992 observations
 - Potential **Analysis** Variables:
 - ITM QTY
 - Number of items ordered
 - ITM PRICE
 - Price (Cost) of Item

- Task:
 - Obtain the sum of
 - · items sold
 - item prices
 - By
 - MAILCODE
 - STATUS

```
title 'Changes and Enhancements to PROCs MEANS & SUMMARY';

title2 'In Version 8 of the SAS(r) System;

title3 'Simple Example of Using PROC MEANS';

Proc Means NOPRINT

data=orders.orderfile2(where=(milcode in('A','B','C')));

CLASS Milcode Status;

var itm_qty itmprice;

var itm_qty itmprice;

Note: longer variable and data set names in V8

run;

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplicated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration
```

In Version 8 of the SAS(r) System								
Si -p l	e Example of	Using PM	C MEANS					
					total_	tetal_		
Obs	milcode	status	_TYPE_	_FREQ_	sold	revenue		
1		-	•	127742	159889	4250086		
2		1	1	100505	1257 08	3346654		
3		2	1	27237	34181	909432. 4		
4	A	-	2	48792	61385	1628543		
5	B	•	2	23063	28611	763298. 1		
6	C	-	2	55887	69893	1858245		
7	A	1	3	38373	48243	1276468		
8	A	2	3	10419	13142	352074. 8		
9	B	1	3	18073	22430	599346 . 1		
10	B	2	3	4990	6181	163952. 0		
11	C	1	3	44059	55035	1464840		
12	C	2	3	11828	14858	393405. 6		

- What Happened?
 - An output (temporary) SAS Data Set was created which contained analyses of the analysis variables at every possible combination (or "crossing") of the classification variables
 - _TYPE_ variable: 'level' of the aggregation
 - _FREQ_ variable: number of observations in the input, or source, data set which contributed to that row/observation in the output data set

- What Happened? (continued)
 - The output SAS Data Set was printed to the output window
 - Results shown on prior page
 - Can we do a better job of portraying the SAS-generated output?
 - Yes...if we use the Output Delivery System (ODS)

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- · Using the Output Delivery System
 - Can be used to render SAS-generated output in to
 - HTML
 - RTF (new!)
 - Files
 - Here is an example of using ODS with PROC MEANS

```
proc template;
    define style Styles. Andrew/ / store = SASUSER. TEMPLAT;
         parent = styles.miniml;
         style data from cell /
             font = ("arial", 3, Bold)
             background = very light yellow vivid
             foreground = red;
            style table from output /
             borderwidth = 3
             bordercolor = Red:
         style Header from headerandfooters /
             font = ("arial", 3, Bold);
            style systemtitle from titlesandfooters /
             font = ("arial, helvetica, helv", 4, Bold Italic);
            style rowheader from header /
             font = ("Arial, Helvetica, Helv", 3, Bold Italic);
run;
         Note: This document copyright ® 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplicated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration
```

```
ods html
                path= 'c: \' (wrl=none)
                body =' means1. html
                style = styles. andrew7;
 title 'Changes and Enhancements to PROCs MEANS & SUMMANY';
 title2 'In Version 8 of the SAS(r) System;
 title3 'Simple Example of Using PROC MEANS';
 Proc Mass NOPRINT
 data=orders.orderfile2(\text{"here}=(\text{milcode in('A', 'B', 'C')));
 CLASS Mailcode Status;
 var itm_qty itmprice;
 CUTFUT out = new sum=total_sold total_revenue;
 proc print data=new
(ods html close;
 ods listing,
          Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. ® holdcates USA registration
```

Changes and Enhancements to PROCs MEANS & SUMMARY In Version 8 of the SAS(r) System Simple Example of Using PROC MEANS

Obs	mailcode	status	_TYPE_	_FREQ_	total_sold	total_revenue
1			0	127742	159889	4250086
2		1	1	100505	125708	3340654
3		2	1	27237	34181	909432.4
4	A		2	48792	61385	1628543
5	В		2	23063	28611	763298.1
6	С		2	55887	69893	1858245
7	A	1	3	38373	48243	1276468
8	A	2	3	10419	13142	352074.8
9	В	1	3	18073	22430	599346.1
10	В	2	3	4990	6181	163952.0
11	С	1	3	44059	55035	1464840
12	С	2	3	11828	14858	393405.6

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Extensive enhancements in Version 8
 - Improved analytic capabilities
 - New ways to request analyses at different 'crossings' or 'interactions' of the classification variables
 - New features to test the _TYPE_ variable's values
 - Autonaming of variables
 - Multiple CLASS statements permitted

- Extensive Enhancements in V8 (con't.)
 - The **DESCENDING** Option can now be used in the CLASS Statement
 - Multiple CLASS Statements are permitted
 - Multiple VAR Statements are permitted

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Extensive Enhancements in V8 (con't.)
 - The New TYPES Statement
 - Permits user-specified combinations of some of the variables in the CLASS Statement to generate sub-groups/analyses of the numeric variables in the VAR Statement
 - The New WAYS Statement
 - Defines the number ways to combine the CLASS variables
 - Additional Functionalities in the OUTPUT Statement

Improved Analytic Capabilities

- PROCs MEANS/SUMMARY now permits calculation of quantile statistics, such as the
 - Median (50th percentile), 90th percentile, interquartile range
 - Prior to V8, only PROC UNIVARIATE had this capability
 - (PROCs TABULATE and REPORT can also calculate quantile statistics in V8)

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration

Improved Analytic Capabilities

- Task:
 - Obtain the
 - Total revenue (sum)
 - Average revenue (mean)
 - Median revenue (50th percentile)
 - From the Customer Electric Rate Data Set
 - By Rate Schedule

Customer Electric Rate Data Set

- Approximately 16,000 observations
 - Premise-level data
 - Monthly KwH (Kilowatt Hour) Consumption
 - Monthly Billed Revenue
 - Total Revenue
 - Quarterly Revenue
 - Rate Schedule
 - Transformer
 - Region
 - Local Office

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Improved Analytic Capabilities

```
options mocenter modate monumber;

ods html

path= 'c: \' (wrl=none)

body =' wans3. html

style = styles. andrew8;

proc MANS NUPRINT

data=electric.elec_V8;

class rate_schedule;

var total_revenue;

output out=new2 sum=median_REV wean=total_REV

p50=mean_REV;

run;

proc print data=new2;

run;

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration
```

Improved Analytic Capabilities

The SAS System

What's "Wrong" With These Results?

Improved Analytic Capabilities: Quantile Statistics in Version 8

Obs	rate_schedule	_TYPE_	_FREQ_	median_REV	total_REV	mean_REV
1		0	15847	11326177.68	714.72	590.51
2	E1	1	14466	10227564.97	707.01	602.07
3	E1L	1	1225	622299.86	508.00	440.65
4	E1M	1	156	476312.85	3053.29	1040.97

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. ® indicates USA registration

Improved Analytic Capabilities

- · What happened?
 - The desired analyses were carried out, and the data set was created, BUT the variable names are incorrect.
 - This is a common problem when creating new variables in the OUTPUT Statement in PROCs MEANS and SUMMARY
 - Solution: the **AUTONAME** option

The AUTONAME Option

- This option is used in the OUTPUT Statement to have the SAS System automatically name the variables it creates while making an output SAS data set
- The 'analytic operation' is appended to the variable name

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Instit

```
The AUTONAME Option

options nocenter nodate nonumber;

ods html path= 'c:\' (url=none)body ='means4.html

style = styles.andrew8;

proc MEANS NOPRINT

data=electric.elec_V8;

class rate_schedule;

var total_revenue;

output out=new8 sum= mean= p50= /autoname;

run;

proc print data=new8;

title3 'Improved Analytic Capabilities: Quantile Statistics';

title4 'Using the AUTONAME Option';

run;

ods html close;

ods html close;
ods 11%phs6/riten consent of the copyright bolder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration
```

The AUTONAME Option

Changes and Enhancements to PROCs MEANS & SUMMARY In Version 8 of the SAS(r) System Improved Analytic Capabilities: Quantile Statistics Using the AUTONAME Option

Obs	rate_schedule	_TYPE_	_FREQ_	total_revenue_Sum	total_revenue_Mean	total_revenue_P50
1		0	15847	11326177.68	714.72	590.51
2	E1	1	14466	10227564.97	707.01	602.07
3	E1L	1	1225	622299.86	508.00	440.65
4	E1M	1	156	476312.85	3053.29	1040.97

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. @ indicates USA registration

Creating Multiple Output SAS Data Sets in PROCs MEANS/SUMMARY

- An often overlooked, but very powerful, feature of these procedures is that multiple SAS data sets can be created in a single invocation of the procedure.
- This is often very useful (and more efficient) when multiple output data sets, continuing analyses at different combinations of the CLASS variables, are required.

- A review of the _TYPE_ variable
 - Represents the level of aggregation among the variables in the CLASS statement.
 - If there are four CLASS variables, the total number of unique combinations of the variables is 16
 - $2^N = 2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16$ (zero to 15)
 - _TYPE_ can be used to select just a few analyses from all possible combinations of the CLASS variables.
 - In Version 6 you needed to know the desired value of _TYPE_ or create a big data set and then use the bit-testing facility in a Data Step

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. ® indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Example:
 - Create a 'lightly summarized' data set from the order file. The analysis variables of interest are the sum of
 - ITM_QTY (Quantity Ordered)
 - ITMPRICE (Item Price)

proc means noprint data=order.orderfile2;
class milcode dept_nbr segment status;
var ituprice itm_qty;
output out=new sum=;
run;
Observations

Note: This document copyright @ 2000 by Sizes Information Scales; Id. All office and the sex written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

 The resulting data set contains the sum of ITMPRICE and ITM_QTY at every possible combination of the four variables in the CLASS Statement

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Out of the 15 different combinations of the CLASS variables, suppose a subset data set containing the analysis of ITM_QTY and ITMPRICE at each combination of MAILCODE and STATUS is desired
 - Version 6 approach 1: figure out the appropriate value of _TYPE_ and use it in a WHERE Clause
 TYPE = 9
 - Version 6 Approach 2: Use the bit-testing facility in a data step

```
proc means noprint data=order.orderfile2;
class milcode dept_nbr segment status;
var itmprice itm_qty;
output out=new sum=;

Output out=new2(where=(_TYPE_ = 9)) sum=;
run;

Data Set 'new' contains a 'complete analysis' of the
```

Data Set 'new' contains a 'complete analysis' of the variables at all possible combinations of the classification variables, while Data Set 'new2' contains only those observations where _TYPE_ = 9

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

```
data subset;
  set new;
  if _type_= '1001'b;
  run;
```

What are the drawbacks to this approach?

Using the Bit Testing Facility in the Data Step

Version 6 Approach 2

- In Version 8, the CHARTYPES option simplifies creation of multiple output SAS data sets from a single use of PROC MEANS.
- This option converts the numeric values of _TYPE_ to a character variable indicating the combination of CLASS Statement variables
 - Although this variable is composed only of zeros and ones, it is a *character* variable

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. ® indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

 With the CHARTYPES Option, making subset data sets with PROC MEANS in Version 8 is very simple!

```
proc means noprint data=order.orderfile2 chartypes;
class mailcode dept_nbr segment status;
var itmprice itm_qty;
output out=one(where=(_type_ = '1001')) sum=;
output out=two(where=(_type_ = '1011')) sum=;
output out=three(where=(_type_ = '0110')) sum = ;
run;
```

- Another enhancement to PROC MEANS is the TYPES Statement (not to be confused with the variable TYPE)
 - This statement permits creation of specified analyses among the CLASS variables
 - If you have many class variables and only some analyses are required, this approach might be preferable

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. ® indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Example: Use the TYPES Statement with the catalog order file to analyze ITMPRICE and ITM_QTY at the following levels of aggregation:
 - Overall ($_$ TYPE $_$ = 0)
 - Segment by Status (_TYPE_ = 3)
 - Mailcode by Segment (_TYPE_ = 10)
 - Mailcode by Dept_Nbr by Segment (_TYPE_ = 14)

```
proc means noprint data=order.orderfile2;

class mailcode dept_nbr segment status;

types ()

segment * status

mailcode * segment

mailcode * dept_nbr * segment;

var itmprice itm_qty;

output out=a sum = ;

run;

TYPES () requests the overall total

__TYPE__ = 0

Note: This document copyright © 2000 by Sierra Information Servizes, Inc. All rights reserved. This document may not be duplicated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration
```

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Using the **DESCENDING** option and multiple **CLASS** statements
 - In Version 6, values of variables placed in the CLASS statement were portrayed in "sort order"
 - The new **DESCENDING** option overrides this default
 - Multiple CLASS statements are allowed in a single use of PROC MEANS

- Using the new IDGROUP option
 - This option allows output of new variables to a data set created by PROC MEANS.
 - Combines the ID and IDMIN options in the PROC MEANS statement and the MAXID and MINID options in the OUTPUT Statement
 - Allows identification of extreme values of the analysis variables
 - The **OUT[N]** option controls the number of extreme observations to be output to the new data set.

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. ® indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

- Tasks:
- Using the Electric Consumption Data Set
 - Perform the analysis in descending order of the values of REGION
 - Obtain the mean and sum of total revenue
 - Obtain the two largest and two smallest values of total revenue
 - -Output to a new SAS data set

```
ods html path= 'c:\' (url=none) body ='means4. html
style = styles. andrew9;
proc means
data=electric.elec_v8(
where=(transformer in('R2448Y','A4356C','B2348X','D8976V')))
noprint nway;
class region/descending;
class transformer;
var total_revenue;
output out=c(remain=(_freq_ = Customer_Count) drop=_type_)
idgroup (mx(total_revenue) out[2] (total_revenue)=maxrev)
idgroup (min(total_revenue) out[2] (total_revenue)=minrev)

su= __automain;

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the
express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS institute, Inc. in the USA
and other countries. @ indicates USA registration
```

```
proc print data=c split = '_';

title 'Changes & Enhancements to PROCs MEANS &
SUMARY';

title2 'in Version 8 of the SAS System;

title3 'Using the DESCENDING and IDCROUP Options';

run;

ods html close;

ods listing;

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written corsent of the copyright holder: SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration
```

More on the IDGROUP statement

idgroup (mx(total_revenue) out[2] (total_revenue)=mxrev)

output the two largest values of TOTAL_REVENUE at each combination of the values of the CLASS variables. The variable prefixes are MAXREV. The variable names are MAXREV1 and MAXREV2

idgroup (min(total_revenue) out[2] (total_revenue)=minrev)

output the two smallest values of TOTAL_REVENUE at each combination of the values of the CLASS variables. The variable prefixes are MINREV. The variable names are MINREV1 and MINREV2

Note: This document copyright © 2000 by Sierra Information Sarvies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration

Changes & Enhancements to PROCs MEANS & SUMMARY in Version 8 of the SAS System Using the DESCENDING and IDGROUP Options

Obs	REGION	transformer	Customer Count	maxrev 1	maxrev 2	minrev 1	minrev 2	revenue Sum	revenue Mean
1	WESTERN	A4356C	493	2482.42	2482.42	54.75	54.75	337698.48	684.99
2	WESTERN	B2348X	510	2671.17	2671.17	60.00	60.00	325224.96	637.70
3	WESTERN	D8976V	207	2205.09	2205.09	60.00	60.00	153963.98	743.79
4	WESTERN	R2448Y	301	4943.85	4943.85	113.44	113.44	190438.76	632.69
5	SOUTHERN	A4356C	72	2433.38	2433.38	148.42	148.42	73709.35	1023.74
6	SOUTHERN	B2348X	52	1496.53	1496.53	228.80	228.80	36504.67	702.01
7	SOUTHERN	D8976V	12	877.01	877.01	239.06	239.06	7881.57	656.80
8	SOUTHERN	R2448Y	47	1599.85	1599.85	206.38	206.38	32027.50	681.44
9	NORTHERN	A4356C	581	5850.54	5850.54	60.00	60.00	468485.68	806.34
10	NORTHERN	B2348X	490	2722.55	2722.55	130.08	130.08	427256.15	871.95
11	NORTHERN	D8976V	237	2274.44	2274.44	198.86	198.86	189297.28	798.72
12	NORTHERN	R2448Y	332	3158.60	3158.60	232.31	232.31	258361.37	778.20
13	EASTERN	A4356C	481	3101.83	3101.83	60.00	60.00	294379.91	612.02
14	EASTERN	B2348X	432	16152.28	16152.28	50.00	50.00	342410.95	792.62
15	EASTERN	D8976V	182	2683.75	2683.75	88.59	88.59	113840.69	625.50
16	EASTERN	R2448Y	339	1883.67	1883.67	60.00	60.00	198000.63	584.07

A few enhancements to PROC PRINT

```
- N = 'text'
- OBS = 'text'
```

 Allow customization of the output to further enhance its visual quality

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS/SUMMARY in Version 8 of the SAS System

```
proc print data=c
obs = 'Row Number'

N = 'Number of Rows in This Region:'
    'Total Rows in This Report: '
        split = ' ';
by DESCENDING Region;
sum_numeric_;
```

```
label customer_count = 'Customer Count'

total_revenue_sum = 'Total Revenue'

total_revenue_mean = 'Mean Revenue'

mxrev_1 = 'Largest Revenue'

mxrev_2 = '2nd Largest Revenue'

mirrev_1 = 'Smallest Revenue'

mirrev_2 = '2nd Smallest Revenue';

format CustomerCount count_2.

surrev_sum surrev_mean

mxrev_1 mxrev_2

mirrev_1 mirrev_2 dollar_12.2;

title 'Summrizing and Reporting Rata Using the SAS(R) System;

title2 'New Features in PROC and PRINT';

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplicated without the express written consent of the copyright holder. SAS and all other SAS Intstitute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration
```

Summarizing and Reporting Data Using the SAS(R) System New Features in PROC and PRINT

REGION=WESTERN

Row Number	transformer	Customer Count	Largest Revenue	2nd Largest Revenue	Smallest Revenue	2nd Smallest Revenue	Total Revenue	Mean Revenue
1	A4356C	493	\$2,482.42	\$2,482.42	\$54.75	\$54.75	337698.48	684.99
2	B2348X	510	\$2,671.17	\$2,671.17	\$60.00	\$60.00	325224.96	637.70
3	D8976V	207	\$2,205.09	\$2,205.09	\$60.00	\$60.00	153963.98	743.79
4	R2448Y	301	\$4,943.85	\$4,943.85	\$113.44	\$113.44	190438.76	632.69
REGION		1511	\$12,302.53	\$12,302.53	\$288.19	\$288.19	1007326.18	2699.16

Number of Rows in This Region:4

REGION=SOUTHERN

Row Number	transformer	Customer Count	Largest Revenue	2nd Largest Revenue	Smallest Revenue	2nd Smallest Revenue	Total Revenue	Mean Revenue
5	A4356C	72	\$2,433.38	\$2,433.38	\$148.42	\$148.42	73709.35	1023.74
6	B2348X	52	\$1,496.53	\$1,496.53	\$228.80	\$228.80	36504.67	702.01
7	D8976V	12	\$877.01	\$877.01	\$239.06	\$239.06	7881.57	656.80
8	R2448Y	47	\$1,599.85	\$1,599.85	\$206.38	\$206.38	32027.50	681.44
REGION		183	\$6,406.77	\$6,406.77	\$822.66	\$822.66	150123.09	3063.99
		Nu	mber of Re	ows in Thi	s Region:	4		

Note: This document copyright ® 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplicated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration

REGION=NORTHERN

Row Number	transformer	Customer Count	Largest Revenue	2nd Largest Revenue	Smallest Revenue	2nd Smallest Revenue	Total Revenue	Mean Revenue
9	A4356C	581	\$5,850.54	\$5,850.54	\$60.00	\$60.00	468485.68	806.34
10	B2348X	490	\$2,722.55	\$2,722.55	\$130.08	\$130.08	427256.15	871.95
11	D8976V	237	\$2,274.44	\$2,274.44	\$198.86	\$198.86	189297.28	798.72
12	R2448Y	332	\$3,158.60	\$3,158.60	\$232.31	\$232.31	258361.37	778.20
REGION		1640	\$14,006.13	\$14,006.13	\$621.25	\$621.25	1343400.48	3255.21

Number of Rows in This Region:4

REGION=EASTERN

Row Number	transformer	Customer Count	Largest Revenue	2nd Largest Revenue	Smallest Revenue	2nd Smallest Revenue	Total Revenue	Mean Revenue
13	A4356C	481	\$3,101.83	\$3,101.83	\$60.00	\$60.00	294379.91	612.02
14	B2348X	432	\$16,152.28	\$16,152.28	\$50.00	\$50.00	342410.95	792.62
15	D8976V	182	\$2,683.75	\$2,683.75	\$88.59	\$88.59	113840.69	625.50
16	R2448Y	339	\$1,883.67	\$1,883.67	\$60.00	\$60.00	198000.63	584.07
REGION		1434	\$23,821.53	\$23,821.53	\$258.59	\$258.59	948632.18	2614.21
		4768	\$56,536.96	\$56,536.96	\$1,990.69	\$1,990.69	3449481.92	11632.56
					- D!	_		

Number of Rows in This Region:4 Total Rows in This Report: 16

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Using the PRELOADFMT and COMPLETETYPES Options

- By default, a format is applied to the values of a classification variable AFTER PROC MEANS has performed the desired analyses
 - If no observations in the input data set have a formatted value of the classification variable, that formatted value is not portrayed in the PROC MEANS-generated output

- The PRELOADFMT Option is placed in the CLASS statement and instructs PROC MEANS to load the format in to memory before conducting the analysis
- The COMPLETETYPES Option, in the PROC MEANS statement, results in having all the formatted values of the classification variable portrayed, even if there are no observations in the input data set have that formatted value.

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, inc. in the USA and other countries. @ indicates USA registration

Using the PRELOADFMT and COMPLETETYPES Options

- Example:
 - From the Customer Order File
 - Analyze sales by Demographic Segment (SEGMENT)
 - First, a format is applied to the one-byte character variable to 'alter the external representation of the value of the variable'

```
proc formt

library = orders;
value $segfint

'A' = '1) Paycheck to Paycheck (A)'

'B' = '2) Bural Betirees (B)'

'C' = '3) Suburban Strivers (C)'

'I' = '4) Soccer Mons & Bads (I)'

'J' = '5) Young Families (J)'

'L' = '6) Urban Pioneers (L)'

'N' = '7) Stable Golden Years (N)'

'O' = '8) Solidly Suburban (O)'

'B' = '9) Dot Com Bandies (R)';

BUN;

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplicated without the express written consent of the copyright holder. SiS and all other SiS institute product names are trademarks of SiS institute, Inc. in the USA and of other countries. Significants USA spiratation.
```

- Now that this format is available, we can use it to improve the visual appeal of a data set created by PROC MEANS which analyzed the customer order file
 - Tasks:
 - Obtain the average number of items ordered
 - Obtain the mean price of items ordered
 - By SEGMENT

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS & SUMMARY in V8 Using the COMPLETETYPES and PRELOADFMT Options

Obs	SEGMENT	_TYPE_	_FREQ_	ITMPRICE_Mean	ITM_QTY_Mean
1		0	335992	\$33.3229	1.24856
2	3) Suburban Strivers (C)	1	268704	\$32.6979	1.24544
3	4) Soccer Moms & Dads (I)	1	10592	\$35.3829	1.23688
4	6) Urban Pioneers (L)	1	22250	\$29.2650	1.20548
5	7) Stable Golden Years (N)	1	32828	\$39.8366	1.23626
6	8) Solidly Suburban (O)	1	1194	\$40.0112	3.17588
7	9) Dot Com Dandies (R)	1	424	\$67.7493	1.29953

- Although there were nine values in the format, only six values of SEGMENT were portrayed in the preceding output
 - That's because no customer records exist in the data set for customers in those three demographic categories.
- By default, PROC MEANS ordered the observations in the output data set using the formatted value of the classification variable.

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Using the PRELOADFMT and COMPLETETYPES Options

- Using the PRELOADFMT and COMPLETETYPES options generates a data set with ALL formatted values of the variable.
- Also, we can re-arrange the ordering of the CLASS variables using the ORDER= option.

- In the next example, the PRELOADFMT and COMPLETETYPES Options are used to generate a data set that includes values of the CLASS variable (demographic segment) for which there are no observations in the analysis data set
- The ORDER=FREQ option in the CLASS statement instructs PROC MEANS to order the observations from highest to lowest frequency

Note: This document copyright @ 2000 by Sierra Information Servies, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Using the PRELOADFMT and COMPLETETYPES Options

Changes and Enhancements to PROCs MEANS & SUMMARY in V8 Using the COMPLETETYPES and PRELOADFMT Options Using the ORDER=FREQ Option

Obs	SEGMENT	_TYPE_	_FREQ_	ITMPRICE_Mean	ITM_QTY_Mean
1		0	335992	\$33.3229	1.24856
2	3) Suburban Strivers (C)	1	268704	\$32.6979	1.24544
3	7) Stable Golden Years (N)	1	32828	\$39.8366	1.23626
4	6) Urban Pioneers (L)	1	22250	\$29.2650	1.20548
5	4) Soccer Moms & Dads (I)	1	10592	\$35.3829	1.23688
6	8) Solidly Suburban (O)	1	1194	\$40.0112	3.17588
7	9) Dot Com Dandies (R)	1	424	\$67.7493	1.29953
8	1) Paycheck to Paycheck (A)	1	0		
9	2) Rural Retirees (B)	1	0		
10	5) Young Families (J)	1	0		

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Another Option in the CLASS Statement

 Another option (shown in a previous example) is ORDER=DESCENDING

Summary and Conclusions

- Significant enhancements to PROCs MEANS and SUMMARY are available to SAS Users in Version 8 of the SAS System.
 - Quantile Statistics
 - TYPES and WAYS Statements
 - CHARTYPE, PRELOADFMT Options
 - IDGROUP Option in the Output Statement

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplicated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration

Summary and Conclusions

- Additional Options Not Discussed in this Presentation
 - CLASSDATA
 - GROUPINTERNAL
 - KEEPLEN
 - NOTRAP
 - INHERIT/NOINHERIT
 - AUTOLABEL
 - Creating _LEVELS_ and _WAYS_ Variables in output data sets
 - Using Multilevel Formats (MLF)

Learning More...

http://www.sas.com/service/library/onlinedoc/v8/whatsnew/tw5508/z1335349.htm

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. © indicates USA registration

Using Multilevel Formats with PROCs MEANS and SUMMARY

- A new feature available in PROC FORMAT allows creation of multilevel formats.
- The next example shows how a multilevel format can be used in PROC MEANS
- The MLF option in the CLASS statement instructs PROC MEANS to create subgroup combinations when a multilabel format is assigned to a classification variable.

```
proc format library = orders;

value totalf (multilabel)

1 - 1000 = '1 to 1,000'

1001 - 2000 = '1,001 to 2,000'

2001 - 3000 = '2,001 to 3,000'

3001 - high = '3,001 or more'

1000 - 1500 = 'a) 1,000 to 1,500'

1501 - 2000 = 'b) 1,501 to 2,000'

2001 - 2500 = 'c) 2,001 to 2,500';

run;

Note: This document copyright © 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder: SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. o indicates USA registration
```

```
ods html path= 'c:\' (url=none) body ='neans4.html
style = styles.andrew8;

proc means noprint completetypes NWff
data=orders.orderhistory;

format tot_sales_ant totalf segment $segfint.;

class segment/preloadfut descending;

class tot_sales_ant/MF preloadfut
order=formatted;

var tot_sales_ant;

output out=new(drop=_type__freq_)

n = Custoner_Count;

run;

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be daplicated without the express written consent of the copyright holder. $AS and all other $AS institute product names are trademarks of $AS institute, Inc. in the USA and other countries. @ indicates USA registration
```

```
proc print data=new noobs;
title1 'Changes and Enhancements to PROCs
MEANS & SUMMARY';
title2 'in Version 8 of the SAS System;
title3 'Using MULTILABEL Forunts';

ods html close;
ods listing;
```

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Changes and Enhancements to PROCs MEANS & SUMMARY in Version 8 of the SAS System Using MULTILABEL Formats

SEGMENT	tot_sales_amt	Customer_Count
9) Dot Com Dandies (R)	1 to 1,000	8
9) Dot Com Dandies (R)	1,001 to 2,000	2
9) Dot Com Dandies (R)	2,001 to 3,000	2
9) Dot Com Dandies (R)	3,001 or more	3
9) Dot Com Dandies (R)	a) 1,000 to 1,500	1
9) Dot Com Dandies (R)	b) 1,501 to 2,000	1
9) Dot Com Dandies (R)	c) 2,001 to 2,500	1
8) Solidly Suburban (O)	1 to 1,000	37
8) Solidly Suburban (O)	1,001 to 2,000	8
8) Solidly Suburban (O)	2,001 to 3,000	2
8) Solidly Suburban (O)	3,001 or more	2
8) Solidly Suburban (O)	a) 1,000 to 1,500	7
8) Solidly Suburban (O)	b) 1,501 to 2,000	1
8) Solidly Suburban (O)	c) 2,001 to 2,500	2

Changes and Enhancements to PROCs MEANS & SUMMARY in Version 8 of the SAS System Using MULTILABEL Formats

Comg mountaine		
5) Young Families (J)	1 to 1,000	0
5) Young Families (J)	1,001 to 2,000	0
5) Young Families (J)	2,001 to 3,000	0
5) Young Families (J)	3,001 or more	0
5) Young Families (J)	a) 1,000 to 1,500	0
5) Young Families (J)	b) 1,501 to 2,000	0
5) Young Families (J)	c) 2,001 to 2,500	0
4) Soccer Moms & Dads (I)	1 to 1,000	332
4) Soccer Moms & Dads (I)	1,001 to 2,000	68
4) Soccer Moms & Dads (I)	2,001 to 3,000	13
4) Soccer Moms & Dads (I)	3,001 or more	14
4) Soccer Moms & Dads (I)	a) 1,000 to 1,500	49
4) Soccer Moms & Dads (I)	b) 1,501 to 2,000	19
4) Soccer Moms & Dads (I)	c) 2,001 to 2,500	9

Note: This document copyright @ 2000 by Sierra Information Services, Inc. All rights reserved. This document may not be duplic ated without the express written consent of the copyright holder. SAS and all other SAS Institute product names are trademarks of SAS Institute, Inc. in the USA and other countries. @ indicates USA registration

Thanks for Inviting Me to SAS Institute UK!